

基于密度估计和VGG-Two的大豆籽粒快速计数方法

王莹¹, 李越¹, 武婷婷², 孙石², 王敏娟^{1*}

(1. 中国农业大学现代精细农业系统集成研究教育部重点实验室, 北京 100083; 2. 中国农业科学院作物科学研究所/农业农村部北京大豆生物学重点实验室, 北京 100081)

摘要: 为快速准确计数大豆籽粒, 提高大豆考种速度和育种水平, 本研究提出了一种基于密度估计和VGG-Two (VGG-T) 的大豆籽粒计数方法。首先针对大豆籽粒计数领域可用图像数据集缺乏的问题, 提出了基于数字图像处理技术的预标注和人工修正标注相结合的快速目标点标注方法, 加快建立带标注的公开可用大豆籽粒图像数据集。其次构建了适用于籽粒图像数据集的VGG-T网络计数模型, 该模型基于VGG16, 结合密度估计方法, 实现从单一视角大豆籽粒图像中准确计数籽粒。最后采用自制的大豆籽粒数据集对VGG-T模型进行测试, 分别对有无数据增强的计数准确性、不同网络的计数性能以及不同测试集的计数准确性进行了对比试验。试验结果表明, 快速目标点标注方法标注37,563个大豆籽粒只需花费197 min, 比普通人工标注节约了1592 min, 减少约96%的人工工作量, 大幅降低时间成本和人工成本; 采用VGG-T模型计数, 其评估指标在原图和补丁(patch)情况下的平均绝对误差分别为0.6和0.2, 均方误差为0.6和0.3, 准确性高于传统图像形态学操作以及ResNet18、ResNet18-T和VGG16网络。在包含不同密度大豆籽粒的测试集中, 误差波动较小, 仍具有优良的计数性能, 同时与人工计数和数粒仪相比, 计数11,350个大豆籽粒分别节省大约2.493 h和0.203 h, 实现大豆籽粒的快速计数任务。

关键词: 卷积神经网络; 籽粒计数; 籽粒图像; 点标注; 密度图; VGG-Two; 育种

中图分类号: TP391.4; TP183

文献标志码: A

文章编号: 202101-SA002

引用格式: 王莹, 李越, 武婷婷, 孙石, 王敏娟. 基于密度估计和VGG-Two的大豆籽粒快速计数方法[J]. 智慧农业(中英文), 2021, 3(4): 111-122.

WANG Ying, LI Yue, WU Tingting, SUN Shi, WANG Minjuan. Fast counting method of soybean seeds based on density estimation and VGG-Two[J]. Smart Agriculture, 2021, 3(4): 111-122. (in Chinese with English abstract)

1 引言

据最新统计数据显示, 2019年中国大豆需求量1.1亿吨, 而国内大豆产量仅1810万吨, 约有9500万吨的产量缺口需通过国际市场弥补^[1]。解决大豆产量不足问题的主要方法是提升大豆育种

水平。目前阻碍育种研究加速的原因之一是无法大规模高通量获取大豆表型性状^[2,3]。百粒重是大豆重要的产量性状, 而测量百粒重的前提就是计算籽粒数量。快速精确的大豆籽粒计数能加快考种速度, 促进大豆育种研究, 进而提升大豆育种水平, 对提升大豆产量具有非常重要的意义。

收稿日期: 2021-01-14 修订日期: 2021-09-16

基金项目: 国家自然科学基金面上项目(31971786), 山东省自然科学基金面上项目(ZR2021MC021)

作者简介: 王莹(1994—), 女, 博士研究生, 研究方向为大豆表型技术。E-mail: b20213080632@cau.edu.cn。

*通信作者: 王敏娟(1984—), 女, 副教授, 研究方向为植物表型组技术。电话: 13601117404。E-mail: minjuan@cau.edu.cn。

在早期阶段,常用的籽粒计数方法是人工籽粒计数,但此操作耗时耗力。同时,肉眼的判断具有很大的偶然性和主观性,长时间计数后不可避免会产生误差,导致计数不准确^[4]。相比人工计数,光电种子数粒仪可轻松避免由于偶然性和主观性产生的误差,结构简单、操作方便,对种子无破坏作用;还能起到“一机多用”的功能。随着研究的深入,光电种子数粒计数误差越来越小,但其普遍存在的不足之处在于价格昂贵,计数速度慢,不利于大规模农业生产自动化的发展^[5]。随着计算机技术的发展和图像信息的普遍化,机器视觉逐渐被科研人员应用到大豆籽粒计数领域,如利用腐蚀膨胀法、分水岭算法^[4,6]、特征点匹配^[7]等基于数字图像处理技术的方法实现籽粒识别和计数。荣斐^[6]针对多种子相互粘连的情况,对图像处理方法进行研究,运用腐蚀膨胀法、面积分配法和分水岭算法,实现对黑豆的分割和快速计数。周洪奎^[4]使用距离变换与分水岭相结合的算法实现粘连区域的分割,提出划线分割算法,并加入多线程以实现算法处理速度的提升。Liu等^[8]确定了图像特征点与谷粒数之间的关系,探索了图像特征点的测量方法,并将其与现有的计数方法进行了相似性和差异性的比较,误差率均低于2%。Tan等^[9]提出了杂交水稻粘连籽粒的精确分割和计数算法,该算法根据分水岭分割算法、改进的角点算法和BP神经网络(Back Propagation Neural Network)分类算法分离和计算粘连谷粒数,与人工计数结果相比,所提方法平均准确率为94.63%。基于传统数字图像处理的籽粒计数方法与人工计数和光电种子数粒仪相比,其计数速度确实有所提升,计数精度也有一定提高,但该类方法需要专业知识和手动提取图像特征,具有复杂的调参过程,同时每个方法都针对具体应用,其泛化能力及鲁棒性较差^[10]。

随着卷积神经网络(Convolutional Neural Networks, CNN)模型^[11]在诸多领域取得非常成功的应用^[12-14],深度学习技术^[15]也得到农业

领域研究人员的认可。与传统数字图像处理技术相比, CNN模型的优势是自动学习和提取有用特征,实现自动化和智能化计数。虽然CNN在大豆籽粒计数领域研究和应用相对较少,但在其他目标计数^[16]方面有相关研究,如Pound等^[17]建立了一个名为ACID(Annotated Crop Image Dataset)的新数据集,提出了一种可以准确定位小麦尖峰和小穗同时准确分类和计数的多任务深度学习方法。Deng等^[18]建立并测试了基于具有特征金字塔网络(Feature Pyramid Networks, FPN)的Faster R-CNN高精度谷物检测模型,用于自动检测和计数每穗粒数,与人工计数谷粒的结果相比,该模型的平均准确率达到99.4%且检测性能不受品种和水分条件的影响。Wu等^[19]开发了线性回归模型和深度学习模型来计算每穗粒数,其计数准确率分别大于96%和99%。Wu等^[20]采用深度学习方法解决传统图像处理算法的局限性,通过构建基于区域的Faster R-CNN模型并运用迁移学习方法,优化了小麦籽粒检测和计数模型,其平均精度为0.91。翟强等^[21]利用具有不同尺寸感受野的CNN和特征注意模块自适应提取多尺度人群特征,结合密度估计方法实现人群计数。

基于CNN的目标计数的实现为大豆籽粒计数提供了新思路。基于图像的目标计数方法可归纳为两大类^[22]:一是基于检测的方法;二是基于回归的方法,其中包括直接回归和密度图回归。由于大豆籽粒图像密度不一、籽粒小,基于小目标检测的方法需要训练检测器来捕获信息,通过检测器检测目标并计算其数量,但是训练检测器比较复杂,计算量较大^[23]。与此同时在深度CNN架构中经过多次下采样后,深层的特征图将会丢失空间信息。而且基于直接回归进行计数的缺点是没有精确的定位,但基于密度回归的方法跳过了艰巨的识别和分类任务,直接生成密度图,学习图像的局部特征和其相应的密度图之间的映射,再根据密度图积分得到目标计数^[24]。因此本研究将密度估计和CNN相结合,根据籽

粒特征构建 VGG-Two (VGG-T) 模型, 进而实现从单一视觉大豆籽粒图像中快速准确识别大豆籽粒数。

2 数据集构建

大规模标注数据的可用性是深度学习在计算机视觉领域取得成功的原因之一。比较成功的神经网络需要大量参数, 参数的正确工作需要大量数据进行训练, 然而目前缺少公开可用带标注的大豆籽粒数据集。因此本研究首先采集并建立了适用于 CNN 的大豆籽粒图像数据集。

2.1 数据采集

选择种植于中国农业科学院作物科学研究所北京顺义基地的大豆样本。基地大棚示意图如图 1 所示, 其中种植区种植各品种大豆。随机选取 5 个种植区域内的小部分区域 (如图 1 深蓝色区域) 的大豆植株, 品种为“中黄 39”, 接着进行收割、摘荚、人工清除污垢等操作, 后续在 2 号大棚的图像采集区 (如图 1 黄色区域) 进行数字成像。

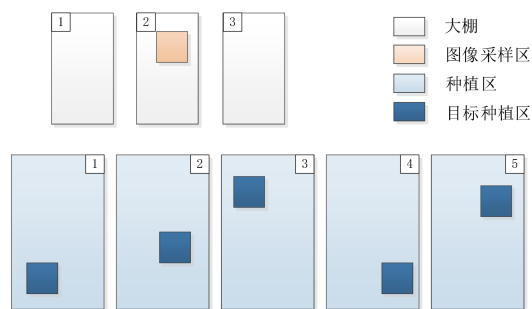


图 1 目标大豆种植区域示意图

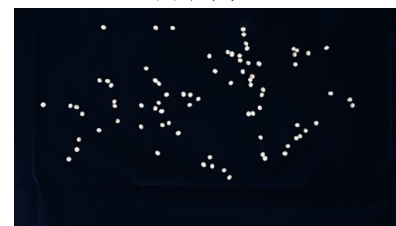
Fig. 1 Sketch map of target soybean planting area

将脱粒大豆种子随机平铺在一块黑色吸光背景布上, 保证种子不重叠, 尽量避免相互接触。在白天、具有漫反射自然照明条件下的植物工厂中, 使用相机 (SONY ILCE-5000 型号, 光圈 $f/4$, 焦距 16 mm, 曝光时间 $1/60$ s, 闪光灯模式为强制无闪光) 采集原始大豆种子图像。采集时, 将相机放置于平铺种子的正上方, 距离种子 30~50 cm。图 2 为大豆籽粒图像的采集装置和经

过调节图像对比度、亮度和尺寸大小等预处理步骤后的原始图像。



(a) 采集装置



(b) 预处理后的原始图像

图 2 大豆籽粒图像采集装置及预处理后的原始图像

Fig. 2 Soybean seed image acquisition device and the original image after pretreatment

2.2 数据标注

2.2.1 大豆籽粒预标注

图 3 是大豆籽粒预标注流程图。为方便后续环节的处理, 运用图像预处理来调节采集的籽粒图像的对比度、亮度、尺寸。针对籽粒图像特点, 利用转灰度、求阈值等一系列图像分析算法进行目标区域提取。为避免错误标注和减少后期人工修正标注的工作量, 补充设计了删除大面积粘连籽粒预标注的处理。最后, 定位部分籽粒并获取这些籽粒质心坐标。

将预处理后的 RGB 籽粒图像 I 转成灰度图 GI , 使用灰度阈值函数计算出全局阈值 $t \in [0, 1]$, 其中灰度阈值函数使用最大类间方差法 (OTSU)。二值化操作是将大于阈值 t 的各像素赋值为 1 (白色), 为目标区域, 其余像素赋值为 0 (黑色), 为背景区域。

为去除籽粒内部的黑色噪声, 对二值图像进行删除小面积处理。如图 4 (a) 红框标注所示。种子内部有黑色区域, 会影响后续对籽粒的识别

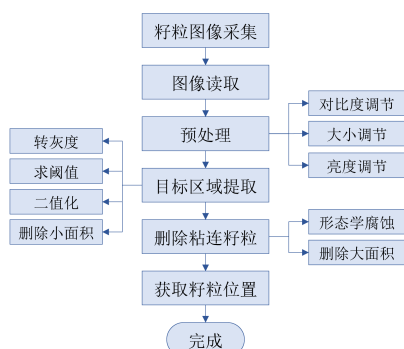


图3 大豆籽粒预标注流程图

Fig. 3 Soybean seed pre-annotation flow chart

和标记,需将籽粒内部全部像素置为1。这里设定面积阈值 T_1 ,将小于 T_1 的区域像素全部置为1,如图4(b)所示。

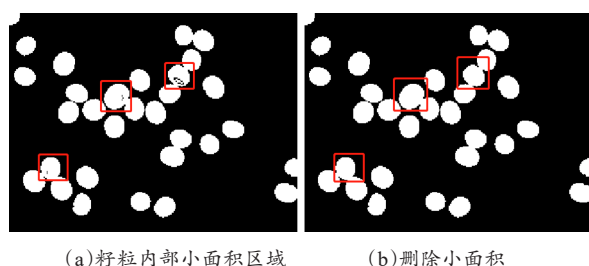


Fig. 4 Sketch map of deleting small area

删除小面积操作之后的二值图像中有部分籽粒粘连,如图5(a)红色框显示,若直接对其进行质心预标注,会出现错标和漏标两个问题,如图6(a)。为尽量减少后期人工修正标注的工作量,利用删除粘连籽粒预标注结果的方法来避免错标问题。图6(b)是删除预标注之后的示意图,只出现漏标情况。这里首先进行形态学腐蚀操作,分离粘连籽粒,如图5(b)红框所示,但存在小部分粘连程度较大的籽粒仍无法分离,如图5(b)绿框所示。为减少后续人工修正错误标注的工作量,进行删除大面积的粘连籽粒预标注处理,其中设置了两个面积阈值 T_2 ,分别是190和300。通过对比质心标注效果,图7(a)为 $T_2=300$ 时的标注示意图,出现了错标和漏标两个问题;图7(b)为 $T_2=190$ 时的标注示意图,只出现了漏标情况,因此将面积阈值参数 T_2 设置

为190。

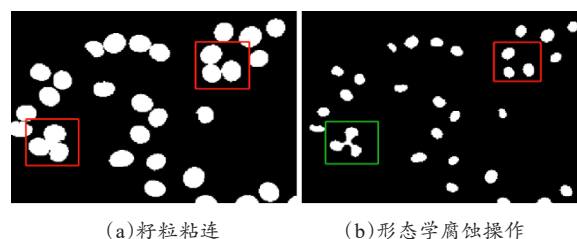


图5 形态学腐蚀操作处理籽粒粘连

Fig. 5 Morphological corrosion treatment of adhesive seeds

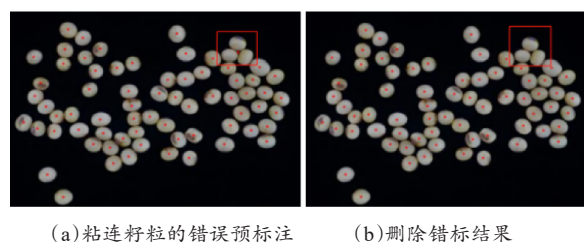


图6 删除粘连籽粒预标注前后的示意图

Fig. 6 Annotation diagram before and after deleting pre-labeling of adhesive seeds

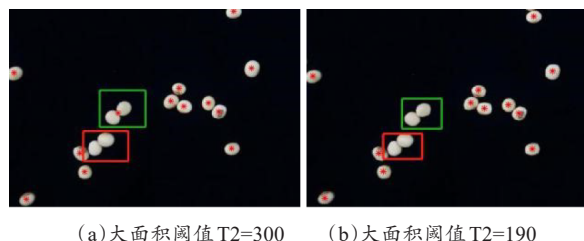


图7 删除不同大面积粘连籽粒的标注示意图

Fig. 7 Annotation diagram after deleting different large area of adhesive seeds

2.2.2 大豆籽粒标注系统

为实现快速、准确、低成本的点标注,利用MATLAB R2017b构建“脱粒种子标注系统V1.0”(简称“标注系统”),其中包括文件管理单元和标注单元两部分。标注系统功能结构如图8所示。文件管理单元用于载入、清空图像,图像中种子所在位置坐标信息的显示和存储等,包括图像载入、参数显示、数据存储以及图像清空功能。标注单元分成两种标注方式:一种是普通的人工标注,即直接在载入的图像上逐一标注种子;另一种是基于传统数字图像处理技术的“预标注+人工修正”的标注方式,首先在MATLAB中调用`imerode()`函数通过形态学腐蚀操作

处理二值化图像，对原始图像进行初步的籽粒识别和定位，详情见 1.2.1，然后在此基础上进行人工补充标注，该方法对 37,563 个大豆籽粒进行标注时只需要使用 197 min，与人工标注相比节约了 1592 min，减少了约 96% 的人工工作量，大大降低标注时间成本和人力成本。

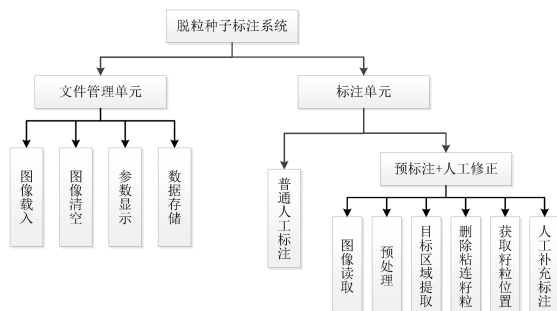


图 8 脱粒种子标注系统功能结构图

Fig. 8 Annotation system function structure diagram of threshed seed

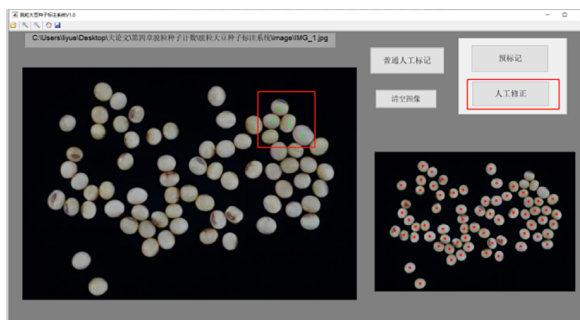
该标注系统的用户界面如图 9 所示。点击图 9 (a) 中“预标记”按钮，红色星号预标记图形将直接显示在大豆籽粒上，从图中可以看出，预标记并没有把所有的籽粒全部标记成功，有部分籽粒被漏标，此时需要人工补充标注。点击图 9 (b) 中“人工修正”按钮，此时左边区域籽粒图像中红色星号标记被清空，右边区域不变，人工参照右边区域的图像预标记情况，通过操作鼠标在左边区域的图像上对没有标记到的籽粒进行补充标注，如图 9 (b)，人工修正利用绿色星号进行标注。标注完成后，点击菜单栏里的“保存”按钮，标注的籽粒坐标被保存为 *.mat 文件。

2.3 数据增强

为在原始图像数量有限的情况下尽可能多的增加输入图像的数量，考虑在每一张原始图像的不同位置裁剪出 9 个补丁 (patch)，patch 的大小设定为原始图像大小的四分之一。设置用于训练和验证网络的图像数量为 239 张，则 patch 数量为 2151 个，远远大于原始图像的数量。在第 4 节



(a) 预标记大豆籽粒



(b) 人工补充标注

图 9 脱粒种子标注系统的标注操作界面

Fig. 9 Annotation operation interface of threshed seed labeling system

试验部分，分别用有无数据增强的数据集对模型进行训练，并进行估测性能的对比，验证了使用数据增强的重要性和必要性。

2.4 数据集建立

按照 6:1:3 的比例，设置训练集、验证集和测试集图像数。大豆籽粒图像的训练集包含 206 张，共 22,582 个标记种子；验证集包含 33 张，共 3631 个标记种子；测试集包含 103 张，共 11,350 个标记种子。经过数据增强，用于训练和验证网络的输入数量扩充为 2151 个 patch。表 1 为该数据集的详细信息。

3 研究方法

3.1 基于密度图的籽粒计数

与基于检测的方法相比较，基于密度图的方法不用进行分类、预选框的回归训练以及目标分割操作，只需要训练网络，将特征图映射成密度

表 1 大豆籽粒图像数据集
Table 1 Soybean seed image dataset

数据集	数量/张	籽粒数/个	最少籽粒 数/个	最多籽粒 数/个	平均籽粒 数/个
训练集	206	22,582	67	201	109
验证集	33	3631	70	200	110
测试集	103	11,350	70	200	110
总数	342	37,563	67	201	109

图即可，然后直接根据密度图积分计算输入图像的籽粒数。

3.2 真值密度图

为实现网络模型从输入种子图像中估测其种子密度图，前期需要对 VGG-T 网络进行训练。训练网络需要提供高质量的训练数据集，基于密度图估计的人群计数通常使用高斯核将标注点生成真值密度图，以真值密度图为监督信号，通过网络生成的密度图计数求和来实现计数，以及计算损失。因此本节所用的数据集除 2.4 小节描述的种子图像外，还包括每张种子图像对应的真值密度图。

将高斯核与种子中心标注点进行卷积操作，可以生成种子图像对应的真值密度图。

$$D^{gt} = \sum_{i=1}^N G_{\sigma}(x - x_i) \quad (1)$$

其中， x_i 表示种子的中心位置； x 表示输入图像中各像素的位置； N 表示该图像包含的种子个数，个； G_{σ} 是高斯核， σ 为扩散参数，值由图像中种子的大小来确定。由于种子相对稀疏并且同一品种的种子大小相差很小，因此在高斯核中使用相同的扩展参数来生成真值密度图。种子图像的真值密度图如图 10 所示。

3.3 VGG-T 网络架构

由于密度回归法多用于人群计数^[25]，且许多工作均采用 VGG16 为主干^[26-28]，以在许多测试数据集上获得良好的性能。根据在不同数据集上的良好表现，本研究同样以 VGG16 作为基础网络。近期的目标计数方法大多使用主干卷积神

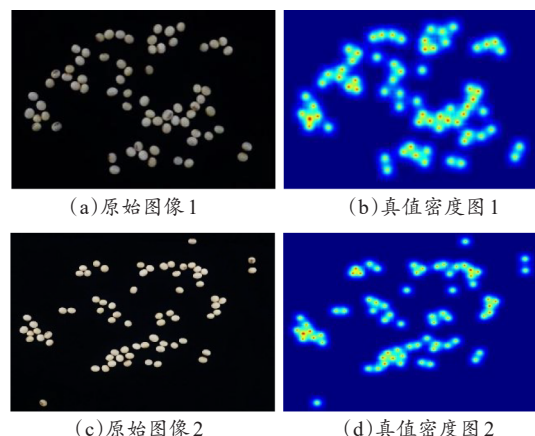


图 10 大豆籽粒真值密度图

Fig. 10 Soybean seed truth density map

经网络的最后一层生成估测密度图。然而由于最后一层卷积层只有单一尺度，同时由于多层池化操作，使得最后一层的分辨率大大降低，不利于生成高精度的密度图，因此这种网络不能实现准确的估测目标数量。为了更好地检测小物体，本研究结合了卷积神经网络浅层和深层的特征，以获得必要的空间和语义信息。

本研究设计的 VGG-T 网络拥有以下特性：在 Conv4_3 之后分支了 2 个特征数据流，第一个数据流直接生成第一个密度图，第二个数据流经过 Conv5_3 之后，生成第二个密度图，将两个密度图进行融合得到最终的估测密度图。图 11 给出了 VGG-T 的架构图。

该网络的主干网络为 VGG16。主要特征提取步骤为：(1) 用 64 个 3×3 、步幅为 1 的滤波器构建的两个卷积层 (Conv1_1, Conv1_2) 对输入图像进行卷积，输出特征图的大小保持不变，通道数量为 64；然后用 2×2 、步幅为 2 的滤波器构建最大池化层 (Pool1)，池化层将输入特征图进行压缩，输出大小是输入大小的 1/2，通道数为 64；(2) 用 128 个 3×3 、步幅为 1 的滤波器构建两个卷积层 (Conv2_1, Conv2_2)，输出特征图的大小保持不变，通道数量为 128；然后用 2×2 、步幅为 2 的滤波器构建最大池化层 (Pool2)，池化层将输入特征图进行压缩，输出大小是输入大小的 1/2，即是原始输入图像的

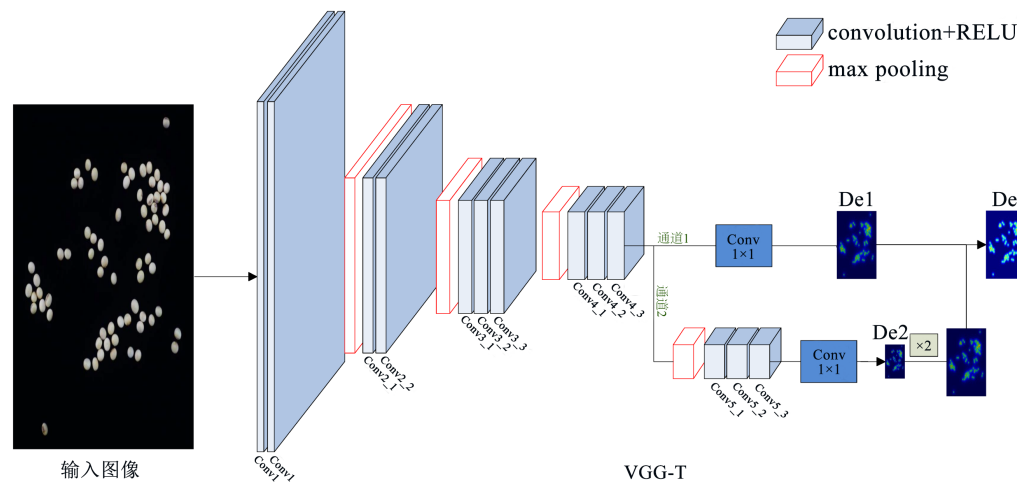


图 11 VGG-T 网络架构图

Fig. 11 Network architecture of VGG-T

1/4, 通道数为 128; (3) 用 256 个 3×3 、步幅为 1 的滤波器构建的三个卷积层 (Conv3_1, Conv3_2, Conv3_3) 对输入图像进行卷积, 输出特征图的大小保持不变, 通道数量为 256; 然后用 2×2 、步幅为 2 的滤波器构建最大池化层 (Pool3), 池化层将输入特征图进行压缩, 输出大小是输入大小的 1/2, 即是原始输入图像的 1/8, 通道数为 256; (4) 用 512 个 3×3 、步幅为 1 的滤波器构建的三个卷积层 (Conv4_1, Conv4_2, Conv4_3) 对输入图像进行卷积, 通道数量为 512。之后分支了 2 个特征数据流。

VGG-T 与传统 VGG16 网络模型的对比如表 2 所示。相比 VGG16, 本研究用 1×1 的卷积核代替全连接层, 因为 1×1 的卷积核一方面大大降低求解网络参数的个数, 同时满足全连接层的作用, 另一方面还能够适应不同的输入数据的大小。其中标 1 的通道直接经滤波器为 1×1 的 Conv 回归得到一个密度图 (De1); 标 2 的通道则需要再经过 Pool4, Conv5_1, Conv5_2, Conv5_3, Conv 回归得到另一个密度图 (De2), 由于比通道 1 多一次池化操作, 其得到的密度图尺寸会再减小一半, 为能够完成最后一步密度图的融合, 该通道还要经过一次反卷积操作。

使用 $D^{e_1} = \{d_j^{e_1}\}$ 和 $D^{e_2} = \{d_j^{e_2}\}$ 分别表示从

表 2 VGG-T 与 VGG 网络结构对比

Table 2 Network structure comparison of VGG-T and VGG

VGG16	VGG-T	
Input (224×224 RGB image)	Input (RGB image)	
Conv1_3 - 64	Conv1_3 - 64	
Conv1_3 - 64	Conv1_3 - 64	
Pool1	Pool1	
Conv2_3 - 128	Conv2_3 - 128	
Conv2_3 - 128	Conv2_3 - 128	
Pool2	Pool2	
Conv3_3 - 256	Conv3_3 - 256	
Conv3_3 - 256	Conv3_3 - 256	
Conv3_3 - 256	Conv3_3 - 256	
Pool3	Pool3	
Conv4_3 - 512	Conv4_3 - 512	
Conv4_3 - 512	Conv4_3 - 512	
Conv4_3 - 512	Conv4_3 - 512	
Pool4	1	2
Conv5_3 - 512	Pool4	
Conv5_3 - 512	Conv5_3 - 512	
Conv5_3 - 512	Conv5_3 - 512	
Pool5	Conv	Conv5_3 - 512
FC - 4096	Conv	
FC - 4096	Deconv	
FC - 1000	Combine	
Soft - max		

Conv4_3 和 Conv5_3 回归得到的两个密度图，由滤波器为 1×1 且只有一个输出的卷积层回归得到。其中， j 表示密度图中第 j 个像素， $d_j^{e_1}$ 表示第 j 个像素的密度。因为经过最大池化操作， D^{e_1} 和 D^{e_2} 有不同的尺寸大小：每经过一个最大池化，输出尺寸都会变成原来的 $1/2$ ， D^{e_1} 经过三次最大池化，其尺寸是输入图像的 $1/8$ ， D^{e_2} 经过四次最大池化，其尺寸是输入图像的 $1/16$ 。相应的，为能够完成网络模型的训练，将真值密度图下采样到原尺寸的 $1/8$ 和 $1/16$ 。使用平均两个估测密度图的方式进行融合：首先，定义 $UP(\cdot)$ 为反卷积上采样过程，使用 $UP(D^{e_2})$ 来表示 D^{e_2} 通过反卷积层进行上采样得到与 D^{e_1} 相同尺寸的密度图；然后使用 $(D^{e_1} + UP(D^{e_2}))/2$ 表示融合这两个相同尺寸的估测密度图，以得到最终的估测密度图 D^e ：

$$D^e = \frac{D^{e_1} + UP(D^{e_2})}{2} \quad (2)$$

其中， D^e 表示融合后的估测密度图代号，其分辨率是输入图像的 $1/8$ ，同时需要下采样相应的真值密度图。

VGG-T 输入的是图像，输出的是种子密度图，对密度图积分可得出该图包含的种子总数，用于计算种子数的公式如下：

$$C^{et}(N) = \sum_{i=0, j=0}^{i=L, j=J} d_{ij}^{et}(N) \quad (3)$$

其中， $C^{et}(N)$ 是测试图像 N 中包含的种子数量估测值，粒； $d_{ij}^{et}(N)$ 表示通过网络最优模型获得的图像 N 的每个像素的估测密度值，粒。

3.4 损失函数计算

均方误差损失函数 L^{MSE} 是典型的损失函数之一，它能逐像素地计算出训练网络中生成的估测密度图与训练数据中给出的真值密度图之间的欧几里德距离，函数如公式 (4) 所示。但是均方误差损失不能考虑到密度图之间的局部相关性，因此使用结构点差异 (Structural Dissimilarity, DSSIM) 损失函数 L^{DSSIM} 来测量估测密度图和真

值密度图之间的局部模式一致性，见公式 (5)。 L^{DSSIM} 源自结构相似性 (Structural Similarity, SSIM)，其函数见公式 (6)。

$$L^{MSE} = \frac{1}{2N} \sum_{i=1}^N \|E(X_i; \Theta) - G_i\|_2^2 \quad (4)$$

$$L^{DSSIM} = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{1}{M} \sum_j SSIM_i(j) \right) \quad (5)$$

$$SSIM_i = \frac{2\mu_{E_i}\mu_{G_i} + C_1}{\mu_{E_i}^2 + \mu_{G_i}^2 + C_1} \times \frac{2\sigma_{E_i G_i} + C_2}{\mu_{E_i}^2 + \mu_{G_i}^2 + C_2} \quad (6)$$

$$L(\theta) = L^{MSE} + \lambda L^{DSSIM} \quad (7)$$

其中， Θ 是在网络中一组可学习的参数； N 为训练图像的数量，个； X_i 表示输入图像； M 是密度图中的像素数，个； λ 是平衡 L^{MSE} 和 L^{DSSIM} 的加权值。 E 和 G 分别表示估测值和真值。 $SSIM_i$ 中的均值 μ_{E_i} 、 μ_{G_i} 和标准差 σ_{E_i} 、 σ_{G_i} 、 $\sigma_{E_i G_i}$ 由大小为 5×5 的高斯滤波器在每个位置 j 上计算得到， $C_1 = (k_1 L)^2$ ， $C_2 = (k_2 L)^2$ 为两个常数，避免除零， $L = 2^B - 1$ 为像素值范围， B 表示比特深度，且 $k_1 = 0.01$ ， $k_2 = 0.03$ 为默认值。方程中忽略了平均值和标准差对像素 j 的依赖性。 $L(\theta)$ 为真值密度图与估测密度图之间的损失。

由于训练样本的数量有限，以及梯度消失对深度神经网络的影响，网络能够同时学习所有参数并不容易。受到预训练的启发，分别对通道 1 和通道 2 单独训练，学习到各层参数作为整体训练时 2 个分支通道的初始值。

3.5 评估指标

使用平均绝对误差 (Mean Absolute Deviation, MAE) 和均方误差 (Mean-Square Error, MSE) 来评估本方法。MAE 是一种常见的用于回归模型的损失函数，反映估测值和真实值之间的距离，定义如下：

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i - a_i| \quad (8)$$

其中， N 为测试样本的数量，个； e_i 为被评

估的模型估测的第 i 张图像中的种子数, 个; a_i 为来自被标记的第 i 张图像中的实际种子数, 个; MAE 表示测试集中种子数估测的准确性, MAE 越小, 说明种子数估测的越准确。

MSE 是最常用的回归损失函数, 表示种子数估测的稳定性, MSE 越大, 说明估测的结果存在异常值。 MSE 定义如下:

$$MSE = \frac{1}{N} \sum_{i=1}^N (e_i - a_i)^2 \quad (9)$$

4 试验与结果分析

试验在操作系统为 Ubuntu 18.4 64-bits 的 PC 机上进行, 其处理器为 Intel® Xeon (R) CPU E5-2630 v4 @ 2.20GHz×20, 内存为 32 GB。使用 PyTorch 深度学习框架基于 NVIDIA 1080Ti GPU 来实现网络训练和测试。

4.1 有无数据增强计数对比

分别使用 239 张大豆籽粒图像 (无数据增强) 和 2151 个 patch (有数据增强) 作为训练数据来训练 VGG-T 网络, 在训练网络的过程中, 使用验证集来评估和优化模型, 然后用分别得到的最优模型来估测测试集中图像的籽粒数。表 3 为相关结果数据。

表 3 VGG-T 在有无数据增强下的试验结果
Table 3 Test results of VGG-T with/without data enhancement

模型	Patch	$MAE/(\text{粒} \cdot \text{图}^{-1})$	$MSE/(\text{粒}^2 \cdot \text{图}^{-1})$
VGG-T	无	0.6	0.6
	有	0.2	0.3

从结果数据可以看出, 使用数据增强的方法生成的 patch 作为训练数据, 其测试后得出的 MAE 和 MSE 数值都较小, 由此证明数据增强这一步骤对于提高网络的估测性能非常重要而且必要。这里使用的随即裁剪增加了训练样本的多样性, 相当于建立每个因子特征与相应类别的权重关系, 减弱背景 (或噪声) 因子的权重, 且使模型面对缺失值不敏感, 最终产生更好的学习效

果, 增加了模型的稳定性, 进而提高网络的估测性能。

4.2 不同计数方法对比

使用 239 张原始大豆籽粒图像和 2151 个 patch 作为训练集来分别训练基础 VGG16 网络、VGG-T 网络、ResNet18 网络和 ResNet18-T 网络。其中 ResNet18-T 采用与 VGG-T 相同的思想增加一个分支以融合多尺度特征。同时使用传统形态学操作, 该过程首先利用最大类间方差法对已进行亮度、大小、对比度调节的图像二值化, 之后通过形态学梯度 (膨胀-腐蚀) 获得大豆籽粒边缘, 接着利用原始图像减去边缘来减少籽粒之间的粘连, 最终使用 OpenCV 中的 findContours 函数找到图像中所有大豆籽粒的轮廓并进行计数, 但从图 5 能够看出经形态学腐蚀后粘连籽粒仍然未能分离。表 4 为 5 种方法的计数性能比较, 根据 MAE 和 MSE 的值可以看出, 当进行传统形态学操作时, 其对应的误差均较大; 对于原始图像, VGG-T 在估测准确性和稳定性上显著优于基础 VGG16、ResNet18 和 ResNet18-T; 对于增强后的 patch 数据, VGG-T 的性能与 ResNet18-T 相当, 同时优于 VGG16 和 ResNet18。综上所述, 试验结果显示了所提出的两个分支进行数据特征的融合能够进一步实现模型性能的提升, 且 VGG-T 网络在所有数据中的综合表现最好。

表 4 5 种方法的计数性能对比

Table 4 Comparison on the count performance of five methods

Method	Patch	$MAE/(\text{粒} \cdot \text{图}^{-1})$	$MSE/(\text{粒}^2 \cdot \text{图}^{-1})$
传统形态学操作	无	4.36	7.17
ResNet18	无	2.1	2.4
	有	0.7	0.8
VGG16	无	2.2	2.9
	有	0.4	0.6
ResNet18-T	无	0.7	0.8
	有	0.2	0.3
VGG-T	无	0.6	0.6
	有	0.2	0.3

4.3 不同测试集计数对比

将测试集的所有图像按照每张含有籽粒数的大小进行升序排列,然后将排好的103张测试图像分成7组,组1~组7分别包含15、15、15、15、14、14、15张籽粒图。用VGG-T的最优训练模型分别测试这7组测试集,表5为MAE、MSE、真值种子数以及估测种子数。真值种子数表示各组平均每张图像含有的种子数的真实值,估测种子数表示各组平均每张图像含有的种子数的估测值。由表中数据可以看出,组1的MAE和组4的MSE分别达到最小,为0.46和0.52。同时随着图像中大豆籽粒数量的不断增加,组1~组7的MAE和MSE大致呈升高趋势,但增加幅度均较小,其中真值和估值最多相差2粒,说明训练得到的最优模型在包含70~200个籽粒图像上均具有优良的计数性能。

表5 不同测试集的试验结果
Table 5 Results of different test sets

组别	MAE/ (粒·图 ⁻¹)	MSE/ (粒 ² ·图 ⁻¹)	真值种子 数/粒	估值种子 数/粒
组1	0.46	0.53	71	71
组2	0.50	0.59	80	79
组3	0.61	0.57	90	90
组4	0.58	0.52	100	101
组5	0.63	0.55	110	109
组6	0.68	0.65	120	118
组7	0.73	0.63	200	199

4.4 时间成本

目前人工计数是大豆育种者使用最为普遍的计数方法,同时光电种子数粒仪可轻松避免偶然性和主观性导致的误差,因此将本研究方法与光电种子数粒仪、人工计数方法进行计数时间比较,结果如表6所示。在采集原始图像的同时,调研了三位大豆育种工作者三天内计数种子的情况,经统计得出人工计数效率为100粒/80s,即1.25粒/s;光电种子数粒仪的计数速度大约为1000粒/3min,即5.56粒/s。利用本研究方法的

计数效率为116.69粒/s。

本研究建立的数据集中测试集共103张大豆籽粒图,包含11,350粒种子,假设计数效率均不变,不间断人工手动计数需要大约2.52h,光电种子数粒仪则需要0.23h左右,而利用本方法耗时大约0.027h。本方法针对人工计数和数粒仪分别节省了大约2.493h和0.203h,所用时间成本分别是人工手动计数、数粒仪时间成本的1/94和1/9。

表6 11,350粒大豆籽粒计数时间成本对比
Table 6 Cost comparison of 11,350 soybean seed

指标	counting time		
	人工计数	光电种子数粒仪	VGG-T
效率/(粒·s ⁻¹)	1.25	5.56	116.690
总耗时/h	2.52	0.23	0.027

5 结论

本研究提出了一种大豆籽粒快速高精度计数方法,构建VGG-T模型并结合籽粒密度图进行回归,所得结论如下:

(1) 设计了大豆籽粒标注系统,提出了基于数字图像处理技术的预标注和人工修正标注相结合的快速目标点标注方法。新方法标注37,563个大豆籽粒只需要花费197min,比普通人工标注节约了1592min,减少了约96%的人工工作量。

(2) 建立了包含342张已标注大豆籽粒图像,共37,563个中心被标注的公开可用大豆籽粒图像数据集。

(3) 构建了结合密度估计方法的基于VGG-T的大豆籽粒数估测模型,其评估指标在原图和patch情况下的MAE分别为0.6和0.2,MSE为0.6和0.3,相比传统图像形态学操作、ResNet18、ResNet18-T和VGG16网络,本方法提高了大豆籽粒计数的准确性。同时相比人工计数和数粒仪,以0.027h完成测试集中11,350个大豆籽粒的快速计数,分别节省了大约2.493h和0.203h。

参考文献:

- [1] 韩昕儒, 梅旭荣, 李思经, 等. 中国农业产业发展战略前瞻[J]. 智库理论与实践, 2019, 4(6): 2-7.
HAN X, MEI X, LI S, et al. The development strategy of China's agricultural industry[J]. Think Tank Theory & Practice, 2019, 4(6): 2-7.
- [2] ALI A, KHAN S A, EHSANULLAH, et al. Estimation of genetic parameters in soybean for yield and morphological characters[J]. Pakistan Journal of Agriculture, Agricultural Engineering, Veterinary Sciences, 2016, 32(2): 162-168.
- [3] 何进. 不同年代大豆品种籽粒产量差异及其水磷亏缺适应机制[D]. 兰州: 兰州大学, 2016.
HE J. Grain yield difference of soybean varieties in different ages and its adaptation mechanism to water and phosphorus deficiency[D]. Lanzhou: Lanzhou University, 2016.
- [4] 周洪奎. 基于图像处理的水稻考种系统的设计与实现[D]. 成都: 电子科技大学, 2019.
ZHOU H. Design and implementation of rice seed test system based on image processing[D]. Chengdu: University of Electronic Science and Technology, 2019.
- [5] 宋初苏, 华娇, 蓝景针, 等. 转盘斜刮式光电自动数粒仪设计[J]. 农业机械学报, 2011, 42(11): 89-92.
SONG R, HUA J, LAN J, et al. Design of photoelectric automatic particle counting instrument with rotary table[J]. Transactions of the CSAM, 2011, 42(11): 89-92.
- [6] 荣斐. 基于图像处理的作物种子自动计数软件开发[J]. 工业设计, 2011(7): 126-127.
RONG F. Development of crop seed automatic counting software based on image processing[J]. Industrial Design, 2011(7): 126-127.
- [7] 崔亮. 基于机器视觉的农作物种子计数检测系统[D]. 太原: 中北大学, 2016.
CUI L. Crop seed counting detection system based on machine vision[D]. Taiyuan: North China University, 2016.
- [8] LIU T, CHEN W, WANG Y, et al. Rice and wheat grain counting method and software development based on Android system[J]. Computers and Electronics in Agriculture, 2017(141): 302-309.
- [9] TAN S, MA X, MAI Z, et al. Segmentation and counting algorithm for touching hybrid rice grains[J]. Computers and Electronics in Agriculture, 2019(162): 493-504.
- [10] 潘锐, 熊勤学, 张文英. 数字图像技术及其在作物表型研究中的应用研究进展[J]. 长江大学学报(自然科学版), 2016, 13(21): 38-41.
PAN R, XIONG Q, ZHANG W. Digital image technology and its application in crop phenotype research[J]. Journal of Changjiang University, 2016, 13(21): 38-41.
- [11] 章琳, 袁非牛, 张文睿, 等. 全卷积神经网络研究综述[J]. 计算机工程与应用, 2020, 56(1): 25-37.
ZHANG L, YUAN F, ZHANG W, et al. A survey of total convolution neural networks [J]. Computer Engineering and Application, 2020, 56(1): 25-37.
- [12] ALSMIRAT M A, AL-ALEM F, AL-AYYOUB M, et al. Impact of digital fingerprint image quality on the fingerprint recognition accuracy[J]. Multimedia Tools and Applications, 2019, 78(3): 3649-3688.
- [13] MEDEN B, MALLI R C, FABIJAN S, et al. Face deidentification with generative deep neural networks[J]. IET Signal Processing, 2017, 11(9): 1046-1054.
- [14] YU H, HE F, PAN Y. A novel segmentation model for medical images with intensity inhomogeneity based on adaptive perturbation[J]. Multimedia Tools and Applications, 2019, 78(9): 11779-11798.
- [15] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [16] AICH S, STAVNESS I. Global sum pooling: A generalization trick for object counting with small datasets of large images[J/OL]. arXiv:1805.11123. 2018.
- [17] POUND M P, ATKINSON J A, WELLS D M, et al. Deep learning for multi-task plant phenotyping[C]// The IEEE International Conference on Computer Vision Workshops. Piscataway, New York, USA: IEEE, 2017: 2055-2063.
- [18] DENG R, TAO M, HUANG X, et al. Automated counting grains on the rice panicle based on deep learning method[J]. Sensors, 2021, 21(1): 281.
- [19] WU W, LIU T, ZHOU P, et al. Image analysis-based recognition and quantification of grain number per panicle in rice[J]. Plant Methods, 2019, 15: ID 122.
- [20] WU W, YANG T, LI RUI, et al. Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales[J]. Journal of Integrative Agriculture, 2020, 19(8): 1998-2008.
- [21] 翟强, 王陆洋, 殷保群, 等. 基于尺度自适应卷积神经网络的人群计数算法[J]. 计算机工程, 2020, 46(2): 250-254.
ZHAI Q, WANG L, YIN B, et al. Crowd counting algorithm based on scale adaptive convolution neural network[J]. Computer Engineering, 2020, 46(2): 250-254.
- [22] AICH S, STAVNESS I. Improving object counting with heatmap regulation[J/OL]. ArXiv: abs/1803.05494. 2018.
- [23] LIU Y, SUN P, WERGELES N, et al. A survey and performance evaluation of deep learning methods for small object detection[J]. Expert Systems with Applications, 2020, 153: 113695.

- tions, 2021, 172: ID 114602.
- [24] BABU SAM D, SURYA S, VENKATESH BABU R. Switching convolutional neural network for crowd counting[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2017: 5744-5752.
- [25] MA Z, WEI X, HONG X, et al. Bayesian loss for crowd count estimation with point supervision[C]// The IEEE/CVF International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2019: 6142-6151.
- [26] VARIOR R R, SHUAI B, TIGHE J, et al. Multi-scale attention network for crowd counting[J/OL]. arXiv: 1901.06026. 2019.
- [27] ZHU L, ZHAO Z, LU C, et al. Dual path multi-scale fusion networks with attention for crowd counting[J/OL]. arXiv: 1902.01115. 2019.
- [28] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv: 1409.1556. 2014.

Fast Counting Method of Soybean Seeds Based on Density Estimation and VGG-Two

WANG Ying¹, LI Yue¹, WU Tingting², SUN Shi², WANG Minjuan^{1*}

(1. Key Laboratory of Modern Precision Agriculture System Integration Research, China Agricultural University, Beijing 100083, China; 2. Institute of Crop Sciences, Chinese Academy of Agricultural Sciences/Beijing Key Laboratory of Soybean Biology, Ministry of Agriculture and Rural Affairs, Beijing 100081, China)

Abstract: In order to count soybean seeds quickly and accurately, improve the speed of seed test and the level of soybean breeding, a method of soybean seed counting based on VGG-Two (VGG-T) was developed in this research. Firstly, in view of the lack of available image dataset in the field of soybean seed counting, a fast target point labeling method of combining pre-annotation based on digital image processing technology with manual correction annotation was proposed to speed up the establishment of publicly available soybean seed image dataset with annotation. Only 197 min were taken to mark 37,563 seeds when using this method, which saved 1592 min than ordinary manual marking and could reduce 96% of manual workload. At the same time, the dataset in this research is the largest annotated data set for soybean seed counting so far. Secondly, a method that combined the density estimation-based and the convolution neural network (CNN) was developed to accurately estimate the seed count from an individual threshed seed image with a single perspective. Thereinto, a CNN architecture consisting of two columns of the same network structure was used to learn the mapping from the original pixel to the density map. Due to the very limited number of training samples and the effect of vanishing gradients on deep neural networks, it is not easy for the network to learn all parameters at the same time. Inspired by the success of pre-training, this research pre-trained the CNN in each column by directly mapping the output of the fourth convolutional layer to the density map. Then these pre-trained CNNs were used to initialize CNNs in these two columns and fine-tune all parameters. Finally, the model was tested, and the effectiveness of the algorithm through three comparative experiments (with and without data enhancement, VGG16 and VGG-T, multiple sets of test set) was verified, which respectively provided 0.6 and 0.2 mean absolute error (MAE) in the original image and patch cases, while mean squared error (MSE) were 0.6 and 0.3. Compared with traditional image morphology operations, ResNet18, ResNet18-T and VGG16, the method proposed improving the accuracy of soybean seed counting. In the testset containing soybean seeds of different densities, the error fluctuation was small, and it still had excellent counting performance. At the same time, compared with manual counting and photoelectric seed counter, it saved about 2.493 h and 0.203 h respectively for counting 11,350 soybean seeds, realizing rapid soybean seeds counting.

Key words: convolutional neural network; seed counting; seed image; point labeling; density map; VGG-Two; breeding

(登陆 www.smartag.net.cn 免费获取电子版全文)